

0590
1012

BEST AVAILABLE COPY

2

OIPE

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/899,303

DATE: 07/20/2001

TIME: 11:23:22

Input Set : N:\Crf3\RULE60\09899303.txt

Output Set: N:\CRF3\07202001\I899303.raw

SEQUENCE LISTING

ENTERED

4 (1) GENERAL INFORMATION:
6 (i) APPLICANT: MAERTENS, GEERT
7 BOSMAN, FONS
8 DE MARTYNOFF, GUY
9 BUYSE, MARIE-ANGE
11 (ii) TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
12 PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
14 (iii) NUMBER OF SEQUENCES: 111
16 (iv) CORRESPONDENCE ADDRESS:
17 (A) ADDRESSEE: NIXON & VANDERHYE P.C.
18 (B) STREET: 1100 NORTH GLEBE ROAD
19 (C) CITY: ARLINGTON
20 (D) STATE: VIRGINIA
21 (E) COUNTRY: U.S.A.
22 (F) ZIP: 22201-4714
24 (v) COMPUTER READABLE FORM:
25 (A) MEDIUM TYPE: Floppy disk
26 (B) COMPUTER: IBM PC compatible
27 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
28 (D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
32 (vi) CURRENT APPLICATION DATA:
C--> 33 (A) APPLICATION NUMBER: US/09/899,303
C--> 34 (B) FILING DATE: 06-Jul-2001
41 (C) CLASSIFICATION:
38 (vii) PRIOR APPLICATION DATA:
39 (A) APPLICATION NUMBER: US 08/612,973
40 (B) FILING DATE: 11-MAR-1996
43 (viii) ATTORNEY/AGENT INFORMATION:
44 (A) NAME: BYRNE, THOMAS E.
45 (B) REGISTRATION NUMBER: 32,205
46 (C) REFERENCE/DOCKET NUMBER: 1487-10
48 (ix) TELECOMMUNICATION INFORMATION:
49 (A) TELEPHONE: (703) 816-4000
50 (B) TELEFAX: (703) 816-4100
54 (2) INFORMATION FOR SEQ ID NO: 1:
56 (i) SEQUENCE CHARACTERISTICS:
57 (A) LENGTH: 21 base pairs
58 (B) TYPE: nucleic acid
59 (C) STRANDEDNESS: single
60 (D) TOPOLOGY: linear
62 (ii) MOLECULE TYPE: cDNA
64 (iii) HYPOTHETICAL: NO
C--> 66 (iv) ANTI-SENSE: NO
71 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
73 GGCATGCAAG CTTAATTAAT T
75 (2) INFORMATION FOR SEQ ID NO: 2:

21

BEST AVAILABLE COPY

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/899,303

DATE: 07/20/2001

TIME: 11:23:22

Input Set : N:\Crif3\RULE60\09899303.txt

Output Set: N:\CRF3\07202001\I899303.raw

```

77      (i) SEQUENCE CHARACTERISTICS:
78          (A) LENGTH: 68 base pairs
79          (B) TYPE: nucleic acid
80          (C) STRANDEDNESS: single
81          (D) TOPOLOGY: linear
83      (ii) MOLECULE TYPE: cDNA
85      (iii) HYPOTHETICAL: NO
C--> 87      (iv) ANTI-SENSE: NO
91      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
93 CCGGGGAGGC CTGCACGTGA TCGAGGGCAG ACACCATCAC CACCATCACT AATAGTTAAT      60
95 TAACTGCA      68
97 (2) INFORMATION FOR SEQ ID NO: 3:
99      (i) SEQUENCE CHARACTERISTICS:
100          (A) LENGTH: 642 base pairs
101          (B) TYPE: nucleic acid
102          (C) STRANDEDNESS: single
103          (D) TOPOLOGY: linear
105      (ii) MOLECULE TYPE: cDNA
107      (iii) HYPOTHETICAL: NO
C--> 109      (iv) ANTI-SENSE: NO
112      (ix) FEATURE:
113          (A) NAME/KEY: CDS
114          (B) LOCATION: 1..639
116      (ix) FEATURE:
117          (A) NAME/KEY: mat_peptide
118          (B) LOCATION: 1..636
121      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
123 ATG CCC GGT TGC TCT TTC TCT ATC TTC CTC TTG GCT TTA CTG TCC TGT      48
124 Met Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu Ser Cys
125 1 5 10 15
127 CTG ACC ATT CCA GCT TCC GCT TAT GAG GTG CGC AAC GTG TCC GGG ATG      96
128 Leu Thr Ile Pro Ala Ser Ala Tyr Glu Val Arg Asn Val Ser Gly Met
129 20 25 30
131 TAC CAT GTC ACG AAC GAC TGC TCC AAC TCA AGC ATT GTG TAT GAG GCA      144
132 Tyr His Val Thr Asn Asp Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala
133 35 40 45
135 GCG GAC ATG ATC ATG CAC ACC CCC GGG TGC GTG CCC TGC GTT CGG GAG      192
136 Ala Asp Met Ile Met His Thr Pro Gly Cys Val Pro Cys Val Arg Glu
137 50 55 60
139 AAC AAC TCT TCC CGC TGC TGG GTA GCG CTC ACC CCC ACG CTC GCA GCT      240
140 Asn Asn Ser Ser Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala
141 65 70 75 80
143 AGG AAC GCC AGC GTC CCC ACC ACG ACA ATA CGA CGC CAC GTC GAT TTG      288
144 Arg Asn Ala Ser Val Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu
145 85 90 95
147 CTC GTT GGG GCG GCT GCT CTC TGT TCC GCT ATG TAC GTG GGG GAT CTC      336
148 Leu Val Gly Ala Ala Ala Leu Cys Ser Ala Met Tyr Val Gly Asp Leu
149 100 105 110
151 TGC GGA TCT GTC TTC CTC GTC TCC CAG CTG TTC ACC ATC TCG CCT CGC      384

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/899,303

DATE: 07/20/2001

TIME: 11:23:22

Input Set : N:\Crf3\RULE60\098993.txt

Output Set: N:\CRF3\07202001\09899303.raw

```

152 Cys Gly Ser Val Phe Leu Val Ser Glp Leu Phe Thr Ile Ser Pro Arg
153          115          120          125
155 CGG CAT GAG ACG GTG CAG GAC TGC AAT TGC TCA ATC TAT CCC GGC CAC      432
156 Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly His
157      130          135          140
159 ATA ACA GGT CAC CGT ATG GCT TGG GAT ATG ATG ATG AAC TGG TCG CCT      480
160 Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp Ser Pro
161 145          150          155          160
163 ACA ACG GCC CTG GTG GTA TCG CAG CTG CTC CGG ATC CCA CAA GCT GTC      528
164 Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro Gln Ala Val
165          165          170          175
167 GTG GAC ATG GTG GCG GGG GCC CAT TGG GGA GTC CTG GCG GGC CTC GCC      570
168 Val Asp Met Val Ala Gly Ala His Trp Gly Val Ala Gly Leu Ala
169          180          185          190
171 TAC TAT TCC ATG GTG GGG AAC TCG GCT AAG GTT TTG ATT GTG ATG CTA      624
172 Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val Leu Ile Val Met Leu
173      195          200          205
175 CTC TTT GCT CTC TAATAG      642
176 Leu Phe Ala Leu
177      210
180 (2) INFORMATION FOR SEQ ID NO: 4:
182      (i) SEQUENCE CHARACTERISTICS:
183          (A) LENGTH: 212 amino acids
184          (B) TYPE: amino acid
185          (D) TOPOLOGY: linear
187      (ii) MOLECULE TYPE: protein
189      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
191 Met Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu Ser Cys
192 1          5          10          15
194 Leu Thr Ile Pro Ala Ser Ala Tyr Glu Val Arg Asn Val Ser Gly Met
195          20          25          30
197 Tyr His Val Thr Asn Asp Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala
198          35          40          45
200 Ala Asp Met Ile Met His Thr Pro Gly Cys Val Pro Cys Val Arg Glu
201          50          55          60
203 Asn Asn Ser Ser Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala
204 65          70          75          80
206 Arg Asn Ala Ser Val Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu
207          85          90          95
209 Leu Val Gly Ala Ala Ala Leu Cys Ser Ala Met Tyr Val Gly Asp Leu
210          100          105          110
212 Cys Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Ile Ser Pro Arg
213          115          120          125
215 Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly His
216          130          135          140
218 Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp Ser Pro
219 145          150          155          160
221 Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro Gln Ala Val
222          165          170          175

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/899,303

DATE: 07/20/2001

TIME: 11:23:22

Input Set : N:\Crf3\RULE60\09899303.txt

Output Set: N:\CRF3\07202001\I899303.raw

224 Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu Ala Gly Leu Ala
 225 180 185 190
 227 Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val Leu Ile Val Met Leu
 228 195 200 205
 230 Leu Phe Ala Leu
 231 210

234 (2) INFORMATION FOR SEQ ID NO: 5:

236 (i) SEQUENCE CHARACTERISTICS:

237 (A) LENGTH: 795 base pairs

238 (B) TYPE: nucleic acid

239 (C) STRANDEDNESS: single

240 (D) TOPOLOGY: linear

242 (ii) MOLECULE TYPE: cDNA

244 (iii) HYPOTHETICAL: NO

C--> 246 (iv) ANTI-SENSE: NO

249 (ix) FEATURE:

250 (A) NAME/KEY: CDS

251 (B) LOCATION: 1..792

253 (ix) FEATURE:

254 (A) NAME/KEY: mat_peptide

255 (B) LOCATION: 1..789

258 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

260	ATG	TTG	GGT	AAG	GTC	ATC	GAT	ACC	CTT	ACA	TGC	GGC	TTC	GCC	GAC	CTC	48
261	Met	Leu	Gly	Lys	Val	Ile	Asp	Thr	Leu	Thr	Cys	Gly	Phe	Ala	Asp	Leu	
262	1				5					10					15		
264	GTG	GGG	TAC	ATT	CCG	CTC	GTC	GGC	GCC	CCC	CTA	GGG	GGC	GCT	GCC	AGG	96
265	Val	Gly	Tyr	Ile	Pro	Leu	Val	Gly	Ala	Pro	Leu	Gly	Gly	Ala	Ala	Arg	
266			20					25					30				
268	GCC	CTG	GCG	CAT	GGC	GTC	CGG	GTT	CTG	GAG	GAC	GGC	GTG	AAC	TAT	GCA	144
269	Ala	Leu	Ala	His	Gly	Val	Arg	Val	Leu	Glu	Asp	Gly	Val	Asn	Tyr	Ala	
270			35				40						45				
272	ACA	GGG	AAT	TTG	CCC	GGT	TGC	TCT	TTC	TCT	ATC	TTC	CTC	TTG	GCT	TTG	192
273	Thr	Gly	Asn	Leu	Pro	Gly	Cys	Ser	Phe	Ser	Ile	Phe	Leu	Leu	Ala	Leu	
274		50				55				60							
276	CTG	TCC	TGT	CTG	ACC	GTT	CCA	GCT	TCC	GCT	TAT	GAA	GTG	CGC	AAC	GTG	240
277	Leu	Ser	Cys	Leu	Thr	Val	Pro	Ala	Ser	Ala	Tyr	Glu	Val	Arg	Asn	Val	
278	65				70				75				80				
280	TCC	GGG	ATG	TAC	CAT	GTC	ACG	AAC	GAC	TGC	TCC	AAC	TCA	AGC	ATT	GTG	288
281	Ser	Gly	Met	Tyr	His	Val	Thr	Asn	Asp	Cys	Ser	Asn	Ser	Ser	Ile	Val	
282			85					90					95				
284	TAT	GAG	GCA	GCG	GAC	ATG	ATC	ATG	CAC	ACC	CCC	GGG	TGC	GTG	CCC	TGC	336
285	Tyr	Glu	Ala	Ala	Asp	Met	Ile	Met	His	Thr	Pro	Gly	Cys	Val	Pro	Cys	
286			100					105					110				
288	GTT	CGG	GAG	AAC	AAC	TCT	TCC	CGC	TGC	TGG	GTA	GCG	CTC	ACC	CCC	ACG	384
289	Val	Arg	Glu	Asn	Asn	Ser	Ser	Arg	Cys	Trp	Val	Ala	Leu	Thr	Pro	Thr	
290			115				120						125				
292	CTC	GCA	GCT	AGG	AAC	GCC	AGC	GTC	CCC	ACC	ACG	ACA	ATA	CGA	CGC	CAC	432
293	Leu	Ala	Ala	Arg	Asn	Ala	Ser	Val	Pro	Thr	Thr	Ile	Arg	Arg	His		
294		130				135				140							

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/899,303

DATE: 07/20/2001

TIME: 11:23:22

Input Set : N:\Crf3\RULE60\09899303.txt

Output Set: N:\CRF3\07202001\I899303.raw

```

296 GTC GAT TTG CTC GTT GGG GCG GCT GCT TTC TGT TCC GCT ATG TAC GTG      480
297 Val Asp Leu Leu Val Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val
298 145      150      155      160
300 GGG GAC CTC TGC GGA TCT GTC TTC CTC GTC TCC CAG CTG TTC ACC ATC      528
301 Gly Asp Leu Cys Gly Ser Val Phe Leu Val Ser Gln Leu Phe rnr Ile
302      165      170      175
304 TCG CCT CGC CGG CAT GAG ACG GTG CAG GAG TGC AAT TGC TCA ATC TAT      576
305 Ser Pro Arg Arg His Glu Thr Val Cln Asp Cys Asn Cys Ser Ile Tyr
306      180      185      190
308 CCC GGC CAC ATA ACG GGT CAC CGT ATG GCT TGG GAT ATG ATG ATG AAC      624
309 Pro Gly His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn
310      195      200      205
312 TGG TCG CCT ACA ACG GCC CTG GTG GTA TCG CAG CTG CTC CGG ATC CCA      672
313 Trp Ser Pro Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro
314      210      215      220
316 CAA GCT GTC GTG GAC ATG GTG GCG GGG GCC CAT TGG GGA GTC CTG GCG      720
317 Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu Ala
318 225      230      235      240
320 GGT CTC GCC TAC TAT TCC ATG GTG GGG AAC TGG GCT AAG GTT TTG ATT      768
321 Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val Leu Ile
322      245      250      255
324 GTG ATG CTA CTC TTT GCT CCC TAATAG      795
325 Val Met Leu Leu Phe Ala Pro
326      260
329 (2) INFORMATION FOR SEQ ID NO: 6:
331 (i) SEQUENCE CHARACTERISTICS:
332 (A) LENGTH: 263 amino acids
333 (B) TYPE: amino acid
334 (D) TOPOLOGY: linear
336 (ii) MOLECULE TYPE: protein
338 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:
340 Met Leu Gly Lys Val Ile Asp Thr Leu Thr Cys Gly Phe Ala Asp Leu
341 1      5      10      15
343 Val Gly Tyr Ile Pro Leu Val Gly Ala Pro Leu Gly Gly Ala Ala Arg
344      20      25      30
346 Ala Leu Ala His Gly Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala
347      35      40      45
349 Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu
350      50      55      60
352 Leu Ser Cys Leu Thr Val Pro Ala Ser Ala Tyr Glu Val Arg Asn Val
353      65      70      75      80
355 Ser Gly Met Tyr His Val Thr Asn Asp Cys Ser Asn Ser Ser Ile Val
356      85      90      95
358 Tyr Glu Ala Ala Asp Met Ile Met His Thr Pro Gly Cys Val Pro Cys
359      100      105      110
361 Val Arg Glu Asn Asn Ser Ser Arg Cys Trp Val Ala Leu Thr Pro Thr
362      115      120      125
364 Leu Ala Ala Arg Asn Ala Ser Val Pro Thr Thr Thr Ile Arg Arg His
365      130      135      140

```

BEST AVAILABLE COPY

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/899,303

DATE: 07/20/2001

TIME: 11:23:23

Input Set : N:\Crf3\RULE60\09899303.txt

Output Set: N:\CRF3\07202001\I899303.raw

L:33 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
L:34 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]
L:66 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:87 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:109 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:246 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:404 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:472 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7
L:541 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:592 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:9
L:649 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:758 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:896 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:916 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:936 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:956 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:976 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:996 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:1016 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:1088 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:21
L:1160 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:1283 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:1413 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:1550 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:1686 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:1823 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:1843 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:1863 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:2118 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:2310 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:2503 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:2681 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:2861 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:3103 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:3449 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:3665 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:49
L:4729 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:4749 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:4771 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:4792 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:4812 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:4832 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:4852 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:4872 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:4892 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:4912 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:4932 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:4952 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/899,303

DATE: 07/20/2001

TIME: 11:23:23

Input Set : N:\Crf3\RULE60\09899303.txt

Output Set: N:\CRF3\07202001\I899303.raw

L:4972 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:4992 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:5012 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:5032 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:5052 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:5072 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]

BEST AVAILABLE COPY